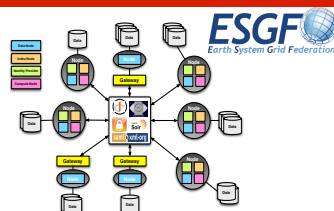# Analytics, Informatics, and Management Systems (AIMS): Software Development

Dean N. Williams on behalf of
Multiple Earth System Communities and Projects
Lawrence Livermore National Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science

CMIP5 Modeling Centers and Federal Data Archive

The **ESGF Peer-To-Peer (P2P)** architecture is based on the concept of a **dynamic system of nodes** that interact on an **egalitarian** basis and can offer a broad range of user and data services, depending on how they are set up. This new system is expressly **designed for extensibility and scalability**, and it supports geospatial and temporal search, a dashboard showing system metrics, user interface for notification, and a rich set of climate analysis tools for data manipulation.
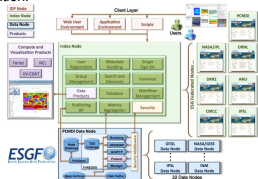
Users can access ESGF data using Web browsers, scripts, and in the near future, client applications. ESGF currently comprises eight national and international portals, four of which hold special status in housing CMIP5/AR5 replication data sets: **LLNL/PCMDI, BADC, DKRZ, and the Australian National University (ANU) National Computational Infrastructure (NCI)**. Users have access to all data from the federation regardless of which ESGF portal is used.
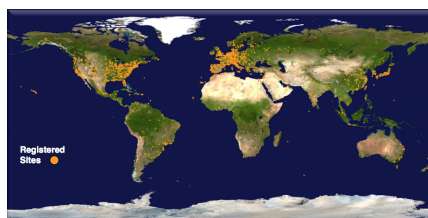

The ESGF user interface portal allows users **access to global collections** of scientific data **from a federation of portals**, as the one shown above on the left. Included on the home page are search capabilities, an information section pertaining to the node and organization, quick links and links to other ESGF P2P nodes, and instructions. The image in the upper right shows the search results for the CMIP5 distributed achieve. The image in the lower right shows a simple plot **Atmospheric InfraRed Sounder (AIRS)** observational data.

Abstract

The objective of the **Analytics, Informatics, and Management Systems (AIMS)** program is to expect and meet today's and tomorrow's demands of the ever growing flood of climate data, analysis and information highways—by creating and delivering the software systems that aim to solve the collaborative and data sharing needs of the climate research community. The AIMS goal is to establish Lawrence Livermore National Laboratory (LLNL) as a leader and visionary architect for all aspects of data discovery and knowledge integration. As a former PCMDI software group, AIMS demonstrated the success of the **Earth System Grid Federation (ESGF)** and the Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT) that comprise an information infrastructure that has revolutionized how climate modeling and intercomparison are performed. Today, the AIMS program leads collaborations to develop an uncertainty quantification diagnostic test bed and data infrastructure for the **Climate Science for a Sustainable Energy Future (CSSEF)** project. Moreover, by working with many others in a global setting, the team is resolving many of the technical challenges regarding scaling and federation (e.g., authentication, sharing, location of data and processing resources, interface standards, etc.) issues that face any attempted large-scale information system. This has been accomplished by using an Open Systems approach that leverages the efforts of a global community.

Diverse Projects as One Data Holdings

Uniting diverse projects in one data holding enables data sharing and intercomparison on the new, unfathomed before scale. Users can apply for group membership to access data from a growing list of climate projects, such as: (a) Phases 3 and 5 of the Coupled Model Intercomparison Project (**CMIP3** and **CMIP5**); (b) Community Climate System Model (**CCSM**); (c) Coordinated Regional climate Downscaling Experiment (CORDEX); (d) Climate Science for a Sustainable Energy Future (**CSSEF**); (e) European Union Cloud Intercomparison, Process Study & Evaluation Project (**EUCLIPSE**); (f) Geo-engineering Model Intercomparison Project (**GeoMIP**); (g) Land-Use and Climate, Identification of robust impacts (**LUCID**); (h) North American Regional Climate Change Assessment Program (**NARCCAP**); (i) Parallel Climate Model (**PCM**); (j) Paleoclimate Modeling Intercomparison Project (**PMIP**); (k) Transpose-Atmospheric Model Intercomparison Project (**TAMIP**); (l) Clouds and Cryosphere (**cloud-cryo**); (m) Observational products more accessible for coupled model intercomparison (**obs4MIPs**); (n) Reanalysis for the coupled model intercomparison (**ANA4MIPs**); and (o) Dynamical Core Model Intercomparison Project (**DCMIP**).

More than 25,000 registered users worldwide

Collaborative Environment & Dashboard

The **ESGF Node Desktop environment**, shown to the right, is a unique, web-based, interactive, graphical user interface for the complete ESGF P2P enterprise system. From this environment, users can operate ESGF web based applications such as:

(1) ESGF Portals: allows federated searching and downloading of data.

(2) **Twitter**: users can exchange ideas live via the integrated Twitter/chart.

(3) **RSS**: users can be notified of data changes

(4) **Whiteboard**: a Web-based collaborative tool that allows multiple users to draw basic forms and objects on the screen.

(5) **Dashboard**: provides a birds eye view of the federation and the ability to drill down into statistical information of the participating systems and their holdings.

(6) **Calendar**: a user scheduler management tool to remember important dates, such as meetings, release dates, and so on.

The Dashboard, shown on the far right, is the perfect interface for node administrators and data providers to see the federation. The Dashboard offers a variety of views on ESGF node and federation statistical information.
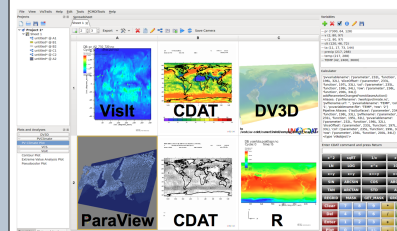
Acknowledgements

New Tools for Analyzing the CMIP Data Archive

UV-CDAT

The Ultra-scale Visualization Climate Data Analysis Tools

Direct client-analysis tool for the ESGF archive, the **UV-CDAT framework** is designed to integrate analytical and visualization tools such as **CDAT**, **VisTrails**, **ParaView**, **VisIt**, **DV3D**, and **R** under one application. Based on Python, it links disparate software subsystems and packages to form an integrated environment for analysis. UV-CDAT's design and openness allow for the collaborative climate community to share development of climate-related software.
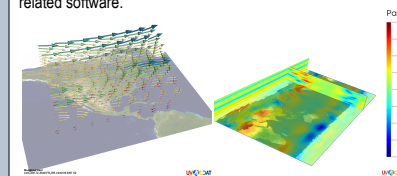
The goals of the UV-CDAT project are to:
(1) develop **derived data** products and **user-reproducible workflows and analysis** archives for the CMIP5 data archive and assessment process;
(2) develop capabilities to **inter-compare ungridded observational data sets** and model data for validation;
(3) deliver efficient **scalable analyses and visualization** for high-resolution simulation data;
(4) deliver data products in formats suitable **for expert and non-expert users**; and
(5) build all capabilities on **existing ESGF P2P node** infrastructures.

The top figure shows the new interactive UV-CDAT graphical user interface (GUI) and the UV-CDAT ESGF P2P node search and browse GUI in the next figure. This allows users to search and browse from UV-CDAT analysis tool as if they were on a Web browser. Once data is located, it is downloaded directly to the UV-CDAT application for manipulation and visualization.

This work is supported by the Regional and Global Climate and Earth System Modeling programs for the Office of Science of the United States Department of Energy. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. LLNL-POST-575614

LLNL Climate SFA Review